



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2015

Voice Äpp: A mobile app for crowdsourcing Swiss German dialect data

Leemann, Adrian ; Kolly, Marie-José ; Goldman, Jean-Philippe ; Dellwo, Volker ; Hove, Ingrid ;
Almajai, Ibrahim ; Grimm, Sarah ; Robert, Sylvain

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: <https://doi.org/10.5167/uzh-114512>
Conference or Workshop Item

Originally published at:

Leemann, Adrian; Kolly, Marie-José; Goldman, Jean-Philippe; Dellwo, Volker; Hove, Ingrid; Almajai, Ibrahim; Grimm, Sarah; Robert, Sylvain (2015). Voice Äpp: A mobile app for crowdsourcing Swiss German dialect data. In: Interspeech 2015, Dresden, 6 September 2015 - 10 September 2015, 2804-2808.



Voice Äpp: a mobile app for crowdsourcing Swiss German dialect data

Adrian Leemann¹, Marie-José Kolly^{2,4}, Jean-Philippe Goldman³, Volker Dellwo⁴, Ingrid Hove⁴, Ibrahim Almajai³, Sarah Grimm⁵, Sylvain Robert⁵, Daniel Wanitsch⁶

¹Department of Theoretical and Applied Linguistics, University of Cambridge

²Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur, LIMSI-CNRS Orsay

³Laboratoire d'Analyse et Technologie du Langage, Université de Genève

⁴Department of Comparative Linguistics, Universität Zürich

⁵Seminar for Statistics, ETH Zürich

⁶iBros.ch

al764@cam.ac.uk, marie-jose.kolly@uzh.ch, jean-philippe.goldman@unige.ch,
volker.dellwo@uzh.ch, ingrid.hove@uzh.ch, ibrahim.almajai@unige.ch,
sarah.grimm@stat.math.ethz.ch, sylvain.robert@stat.math.ethz.ch, daniel@ibros.ch

Abstract

Crowdsourcing speech data through mobile applications is relatively new. In the present contribution we add to the existing body of research an innovative Android and iOS application called 'Voice Äpp'. The free app is pioneering in the sense that it leverages its function as a medium for science communication – thus attracting an extensive user base – to crowdsource audio and dialect data. The app was launched in early 2015 and has already been downloaded 19k times. Nearly half a million audio tokens have been crowdsourced. In this system levels contribution we describe the basic functionalities of the app – voice and dialect analysis –, we present the scientific potential of the corpus created, and discuss methodological issues related to crowdsourcing audio data through mobile applications.

Index Terms: crowdsourcing, Swiss German, automatic speech recognition, dialectology, smartphone application

1. Introduction

The use of crowdsourcing applications to collect speech data has, until recently, received relatively little attention. This is surprising given that most smartphones feature frequency ranges of 50Hz-20kHz [1] and allow for reliable acoustic measurements, e.g. F1 and F2 [2]. A handful of crowdsourcing apps have been developed for the purposes of acoustic modeling, dictionary building, text collection, translation, and dialect mapping: [3, 4] developed Android apps to collect speech for the training of acoustic models. [5, 6] are applications that enable documenting endangered languages. In early 2013, [7] launched the iOS app 'Dialäkt Äpp' (hereafter DÄ) to crowdsource dialect data.

DÄ allows users (1) to do a dialect quiz in which they can determine their own Swiss German (hereafter SwG) dialect and (2) to record their own dialect. DÄ was the number one downloaded free app for iPhones in Switzerland after its release on 22 March, 2013 [8]. It received significant media attention and has >80k downloads [9]. This is not surprising: dialect prediction is a hot topic in linguistics. The New York Times' dialect quiz [10] – based on the Harvard Dialect Survey [11] – which prompts users to indicate their dialectal variants to 25 questions and then tells them where they are from, was the most visited content on the NYT sites in 2013 [12].

Over the past two years we developed an app that builds on DÄ. 'Voice Äpp' (hereafter VÄ) [13, 14] was built as a tool for science communication – of phonetics, dialectology, and linguistics – and crowdsourcing speech. The app was launched on 5 January 2015 and is available for Android [15] and iOS [16]. In the present system levels contribution, we describe the basic functionalities of VÄ, the scientific potential of the data crowdsourced through VÄ, and address methodological considerations of crowdsourcing speech data with mobile applications.

2. Dialäkt Äpp

VÄ is conceptually based on DÄ, which is why we devote the first part of this contribution to DÄ. Audio data crowdsourced through DÄ was used to develop VÄ.

2.1. Dialect prediction

DÄ predicts the users' SwG dialects. The prediction algorithm is based on an overlay of 16 maps from the *Sprachatlas der Deutschen Schweiz* (1962–2003) 'Linguistic Atlas of German-speaking Switzerland' [17, 18]. The Atlas documents regional variation of SwG dialects in the middle of the 20th century. Data was collected in the 1940s in 566 localities. From the perspective of user experience, DÄ prompts its users to select their dialectal variants from a list by tapping on the smartphone screen. After going through the 16 words (all of which may have 2-40 dialectal variants to choose from), DÄ presents users with a list of five localities that best correspond to their set of pronunciation variants. The user can subsequently evaluate the result and – in case of an erroneous prediction – self-declare the dialect s/he actually speaks [9].

2.2. Dialect recordings

The second functionality enables users to record their dialects. Users indicate their dialect, age, and gender before proceeding to the recording screen. They then record the same 16 words as in the prediction function. Until today, DÄ crowdsourced recordings from 3000 speakers, stemming from all parts of German-speaking Switzerland [9, 14].

3. Voice Äpp

VÄ features two core functionalities, which will be presented in more detail here: (1) 'My voice' and (2) 'My dialect'.

3.1. My voice

Upon launching ‘My voice’, the user is prompted to record a sentence in dialect after having entered gender, age, and dialect spoken (Figure 1, left panel). The sentence shown to the speaker (Figure 1, right panel) is taken from the *Gespräch am Neujahrstag*, a standard text used to document dialect variation SwG [19].

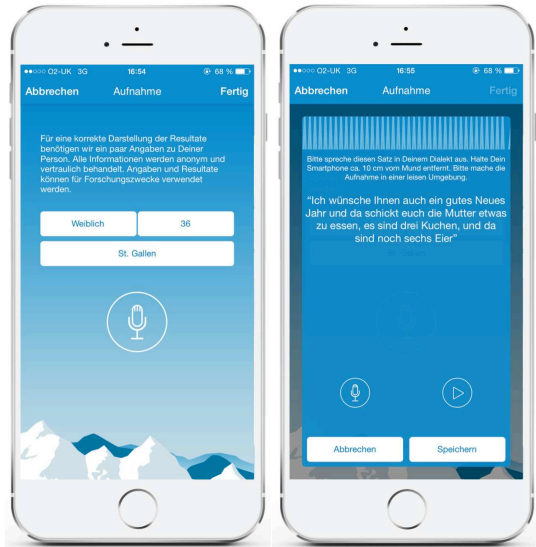


Figure 1: Interface of ‘My voice’.

Once the sentence has been recorded, the user’s speaking rate and pitch are analyzed and displayed. Underlying rate detection is a Praat [20] script that identifies amplitude peaks, where one amplitude peak normally corresponds to one syllable. As a result, the app displays the speech rate in syllables per second. Speech rate is displayed in a histogram that plots all the existing user data, see Figure 2 (left panel).



Figure 2: Interface for speaking rate and f_0 analysis.

The user can compare his speech rate with that of speakers from other dialect regions. In the example shown in Figure 2,

the speaker compares him-/herself to all previously recorded speakers from Bern, who have been reported to be slow speakers based on data crowdsourced through DÄ [21]. Users can further manipulate their speech as ‘faster’ and ‘slower’ (see Figure 2, left panel). f_0 analysis is performed using a Praat [20] script that calculates mean f_0 s. Measurements are displayed in a two-way histogram showing mean f_0 s of all previously recorded males (top) and females (bottom: Figure 2, right panel). Users can also manipulate their f_0 by clicking on the male / female symbol (see Figure 2, right panel), raising or lowering their mean f_0 .

3.2. My dialect

VÄ’s second main functionality determines the user’s dialect through ASR. Much research has been conducted on dialect and accent recognition with the aim of increasing the accuracy of ASR [22, 23, 24, 25]. In our work, however, dialect is identified from the recordings of isolated words. Using HTK [26], we built 15 independent word recognizers with features composed of 13 MFCCs with derivatives at a rate of 100 frames per second. The number of states in word HMMs was either 12 or 14 depending on word length.

Instead of clicking on the corresponding dialect variant (as in DÄ, see 2.1), users speak into the phone their 15 dialectal variants of German words shown on the screen (see Figure 3, left panel, which translates to ‘Dialect word for the core of an eaten apple’). After going through 15 words, VÄ tells the users which dialects they speak by displaying a heatmap (see Figure 3, right panel).

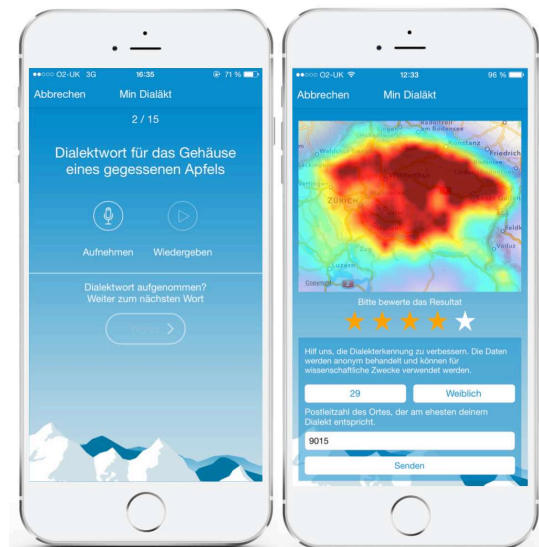


Figure 3: Interface for dialect prediction based on ASR.

To identify dialects, we used information coming both from the historical *Atlas* (i.e., the information used for DÄ’s prediction function) and from more recent data previously crowdsourced through DÄ’s prediction function (i.e., an update of this historical data, cf. 2.1), and combined them in a Bayesian framework. For the following, consider one token and one location. As an informative prior we used a Dirichlet distribution on the variants’ probabilities. The *Atlas* information was incorporated by setting all the concentration on the *Atlas* variant, and giving it a weight equivalent to having observed 30 respondents – one *Atlas* variant for one

location usually means that one to two people have given their variant for the *Atlas*. As this information, which has been collected in the field, may be trusted more than crowdsourced data, however, we gave it more weight. The crowdsourced data (counts per variant) were modeled with a multinomial distribution, whose parameters are the number of respondents and the probabilities of the variants. The posterior distribution, also a Dirichlet distribution, now contains both the *Atlas* information and data that had been crowdsourced through DÄ. Combining the two sources of information in this manner means that we trusted more the data in locations with many respondents, and less locations with less respondents. We used the posterior means as our new variant probabilities per location per token.

The integration of the ASR output proceeded as follows. Given a user's spoken variant, the ASR does not select one variant but rather expresses a level of certainty per variant of the linguistic variable, in the form of a probability distribution. To predict the probability of a location for the heatmap, we computed the average of the posterior means weighted by the ASR probabilities. To combine the answers to all words, we used a naive Bayes estimator (which assumes independence between tokens) with a uniform prior over all locations. The final result was one probability per location, which we displayed with a heatmap by applying a simple kernel regression estimator, see Figure 3 (right panel). Once the users' dialects have been predicted, they can then evaluate the result on a five-point scale (see Figure 3, right panel). They are further prompted to indicate age, gender, and the zip codes that best correspond to their self-declared dialect region.

3.3. Reception of the app

VÄ was launched on 5 January, 2015 and was received well in the public. It was downloaded 19k times (2 March, 2015): 79% of the downloads were for iOS, 21% for Android platforms. Figure 4 shows the download statistics for the first two months. VÄ received considerable media attention in German- and French-speaking Switzerland [27]. 14 users rated Voice Äpp in the App Store, 55 in the Google Play Store with an average rating of 4.5/5, and 3.8/5 points respectively. VÄ was described fun to use and a number of users reported high dialect prediction accuracy. Criticism concerned high battery consumption, slow data transmission, and large amounts of processing power being used when running the app.

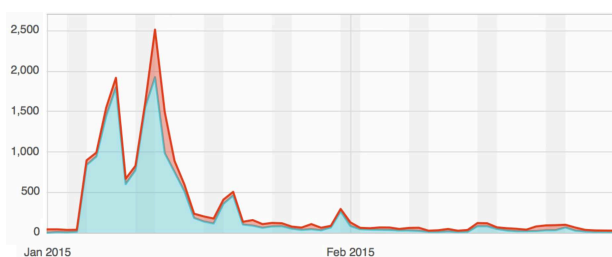


Figure 4: Download statistics of VÄ (blue) and DÄ (red) in the App Store for January and February 2015

4. Use for science

On the one hand, VÄ was primarily developed as a means for science communication. A mobile app as a means of opening a dialogue between phoneticians, linguists, and speech scientists and the public was intentionally selected as we aimed to target a particularly younger audience. On the other hand, VÄ allows

us to crowdsource audio data from the German-speaking population in Switzerland; data which can be used for scientific purposes. Here, we describe in more detail the scientific potential of crowdsourced VÄ data.

4.1. My voice

'My voice' (cf. 3.1) generated 7223 sentence recordings: 5221 of these recordings were from unique users (59% m, 41% f) (for a discussion of multiple submissions, see 5.3.2). Speakers ranged between 17 and 36 years of age (IQR) and mostly spoke dialects from the cantons Zürich, Bern, Aargau or St. Gallen (67%). This data bears significant potential for phonetic analyses: it can be used to investigate the distribution of acoustic characteristics such as formant and fundamental frequencies, durational characteristics of segmental, suprasegmental, and voice quality features in the German-speaking population. Such statistics on large speaker samples are most desirable in the domain of forensic phonetics, where experts compare speech in criminal and suspect recordings. In order to make such comparisons, experts need to know how such features are distributed in the general population. Currently such population data are only scarcely available, for example for speaking rate [28] or f0 [29]. As recordings were collected from all over German-speaking Switzerland, maps of areal variation can be created to display dialect differences in the acoustic parameters present in the corpus [30].

4.2. My dialect

As of 2 March 2015, 'My dialect' (cf. 3.2) generated 489,878 recordings from 18,190 users. 6952 of these users (62% m, 38% f) sent off their evaluations, i.e. prediction feedback, age, gender, and self-declared dialect (cf. Figure 3, right panel). Users were between 16 and 40 years of age (IQR). For 70% of these users, ASR predicted the right canton in the top three hits. This is considerable, given that there is a total of 19 cantons where German is an official language (chance level is thus 5.3% (=100/19)). Moreover, for 65% of these users, one of the three predicted top hits was less than 20 km away from their self-declared dialect. Qualitatively, users rated prediction scores relatively well on the 5 star rating scale (cf. Figure 3, right panel): with a 3.12 average score (SD=1.32). Finally, we intend to perform diachronic analyses with the data collected in this function: given that the words chosen for the prediction in 'My dialect' are based on the *Atlas* – which documented the linguistic situation in German-speaking Switzerland around 1950 – we can now investigate how the dialects have changed over the past 70 years.

5. Methodological considerations

The large volume of speech data that can be crowdsourced with a mobile application is an obvious benefit. Collecting a corpus of this size and geographical spread with traditional dialectological methods would be costly and time-consuming. There are a number of pitfalls to this method of data collection too, nonetheless. Here we address potential limitations of speech data crowdsourced in 'My voice' (5.1) and 'My dialect' (5.2). In 5.3 we address general limitations of speech data that has been crowdsourced with mobile applications. This list of limitations is not exhaustive, of course. In 5.4 we address measures for quality control.

5.1. Limitations of crowdsourced ‘My voice’ data

The quality of recordings crowdsourced in DÄ from 3000 speakers (cf. 2.2) as well as those crowdsourced in VÄ from 5221 speakers (cf. 4.1) is different from the data quality one normally achieves in phonetic studies: recordings are typically conducted in sound-treated rooms. Analyses of the DÄ corpus have shown that around 5% of the recordings were discarded because the speakers were uncooperative; about 5–10% were not usable due to factors such as a high level of background noise. Furthermore, speakers may not always adhere to the instructed 10 cm microphone distance, thus fluctuations in amplitude are common. One upside of these ‘unfavorable’ effects, however, is that these factors increase the validity of the dialect prediction results: the ASR trained on DÄ data seems to function relatively well (cf. 4.2). Results stemming from such a data set are thus more likely to generalize better to real-world situations [31].

5.2. Limitations of crowdsourced ‘My dialect’ data

As mentioned in 4.2, a future aim is to examine how the SwG dialects have changed over the past 70 years based on ‘My dialect’ data. The words chosen for the prediction function stem from the *Atlas*, which documented the state of SwG dialects at around 1950. We can now compare this historical data with how people speak nowadays, as indicated by VÄ user feedback. The methods applied in the *Atlas* are very different from the methods applied here: for the *Atlas*, researchers went into the field and conducted interviews. App data was collected indirectly, with no researcher present. There is much less control over how the data is elicited in the indirect method. In the *Atlas*, it was especially older speakers who had lived in the respective locality for a long time which were carefully selected [32]. In VÄ, a maximum number of SwG speakers were targeted.

5.3. General limitations

5.3.1. Self-declared dialect

The user’s self-declared dialect in ‘My voice’ and ‘My dialect’ is not entirely unproblematic. Here we simply have to assume that the users have an understanding of their linguistic origins. Potentially, users imitated a ‘model’, perhaps more prestigious, dialect when doing the quiz, which would cause the respondents to be a more homogeneous crowd [33] or, nostalgically claimed traditional variants from their communities that they themselves no longer use.

5.3.2. Multiple submissions

In laboratory research, subjects normally only serve once. In web and app-based research, however, there is a distinct possibility for multiple submissions [34]: the same person uses the same app to participate repeatedly or the same person uses the app on different smartphones to participate repeatedly. [35] reported that the rate of repeated participation is below 3% in most studies, however. In VÄ’s ‘My dialect’ prediction recordings, we estimate, however, that 56% of the recordings came about by multiple submissions from the same user: our corpus currently consists of 489,878 single-word-recordings, which was to stem from 32,658 users, if every user articulated the 15 words. When these recordings are grouped by age, gender and device, however, only 18,190 users who recorded

all 15 words remain. This suggests that VÄ users apparently tried to predict their dialect with ASR repeatedly. Perhaps in their second and third attempts they changed their dialect to test their foreign dialect imitation performance? In the ‘My voice’ recording function, too, we suspect multiple recordings by the same user, as we obtained 7223 single sentence recordings which stem from 5221 users (i.e., unique combinations of device, age and gender).

5.3.3. Experimenter bias and trustworthiness

With the method applied, we cannot be sure that users read the instructions in the app. In laboratory settings the researcher can explain in detail the procedures to the participants. This increases chances that subjects understand the instructions. Moreover, the researcher can verify and interact with the participant [35]. Finally, the trustworthiness of participants in online surveys of app-based studies is an often-cited problem [36]. How can we be sure that the users are providing meaningful responses? Proponents of app-based experimentation note, however, that this question applies to all behavioral testing whether app-based or laboratory-based [37].

5.4. Quality control

There are a number of protocols that can be applied to tackle the noisiness in crowdsourced data. For example, a comparison of the performance of the crowd to that of experts [38]. Would experts, i.e. phoneticians or linguists, obtain the same results as laymen if they were to predict their dialect in ‘My dialect’? This validation is difficult to apply, however, as speakers’ linguistic biographies are very diverse. Applying traditional dialectological methods on the phenomena examined, however, enables a validation of results obtained through crowdsourcing [39]. Based on DÄ prediction data (cf. 2.1), we tested whether the spread of /l/-vocalization in SwG as captured by crowdsourced DÄ data would yield similar results as those found with a traditional method. The patterns of diffusion of /l/-vocalization were highly similar [30, 40]. A correlation of the degree of vocalization as captured with the traditional method and the crowdsourcing method was significant at $r=.77$, suggesting a strong, linear relationship between the results obtained with the two methods.

6. Conclusions

The main goal of the current study was to present the basic functionalities of VÄ, describe the scientific potential of the corpus crowdsourced, and to explore methodological issues pertinent to the crowdsourcing framework. One of the significant findings to emerge was that smartphone applications – in particular if presented in a legitimate science communication framework – have great potential of attracting large masses and thus enabling the collection of large volumes of speech and dialect data. Further research might explore how speech parameters such as f_0 and speaking rate are distributed in the corpus crowdsourced through VÄ.

7. Acknowledgments

Warm thanks go out to 65 backers who sponsored DÄ through crowdfunding. We thank the Swiss National Science Foundation for funding ‘Voice Äpp’ (Grant CRA-1_145654).

8. References

- [1] <http://blog.faberacoustical.com/2009/ios/iphone/iphone-microphone-frequency-response-comparison/>, accessed 03 Mar 2015.
- [2] P. De Decker and J. Nycz, "For the Record: Which Digital Media Can be Used for Sociophonetic Analysis?," *University of Pennsylvania Working Papers in Linguistics*, vol. 17, no. 2, pp. 51-59, 2011.
- [3] T. Hughes, K. Nakajima, L. Ha, A. Vasu, P. Moreno and M. LeBeau, "Building transcribed speech corpora quickly and cheaply for many languages," *Proceedings of Interspeech 26.-30.10.2010, Makuhari, Chiba, Japan*, vol. 1914-1917, 2010.
- [4] N. de Vries, M. H. Davel, J. Badenhurst, W. D. Basson, F. de Wet, E. Barnard and A. de Waal, "A smartphone-based ASR data collection tool for under-resourced languages," *Speech Communication*, vol. 56, pp. 119-131, 2014.
- [5] *Ma! Iwaidja*, <https://itunes.apple.com/au/app/ma-iwaidja/id557824618?mt=8>, accessed 03 Mar 2015.
- [6] F. R. Hanke and S. Byrd, "Large-scale text collection for unwritten languages," *International Joint Conference on Natural Language Processing*, pp. 1134-1138, 2013.
- [7] A. Leemann and M.-J. Kolly, *Dialäkt App*, <https://itunes.apple.com/ch/app/dialakt-app/id606559705?mt=8>, accessed 03 Mar 2015.
- [8] <http://www.appannie.com/>, accessed 31 Feb 2014.
- [9] M.-J. Kolly and A. Leemann, "Dialäkt App: Communicating dialectology to the public - crowdsourcing dialects from the public," in A. Leemann, M.-J. Kolly, S. Schmid and V. Dellwo (eds.), *Trends in Phonetics and Phonology. Studies from German-speaking Europe*. Bern: Peter Lang, in press.
- [10] New York Times Dialect Quiz http://www.nytimes.com/interactive/2013/12/20/sunday-review/dialect-quiz-map.html?_r=0, accessed 03 Mar 2014.
- [11] B. Vaux and S. Golder, *The Harvard Dialect Survey*, Cambridge (MA): Harvard Linguistics Department, 2003.
- [12] <http://www.theatlantic.com/technology/archive/2014/01/-em-the-new-york-times-em-most-popular-story-of-2013-was-not-an-article/283167/> accessed 03 Mar 2015.
- [13] J.-Ph. Goldman, A. Leemann, M.-J. Kolly, I. Hove, I. Almajai, V. Dellwo and S. Moran, "A Crowdsourcing Smartphone Application for Swiss German: Putting language documentation in the hands of the users," *LREC*, 2014.
- [14] M.-J. Kolly, A. Leemann, V. Dellwo, J.-P. Goldman, I. Hove and I. Almajai, "Voice App. A smartphone application for crowdsourcing Swiss German dialect data," *Proceedings of Digital Humanities*, 2014.
- [15] <https://play.google.com/store/apps/details?id=ch.voiceapp&hl=en>, accessed 03 Mar 2015.
- [16] <https://itunes.apple.com/ch/app/voice-app/id950037736?mt=8>, accessed 03 Mar 2015.
- [17] *Sprachatlas der deutschen Schweiz*, Bern (I-VI), Basel: Francke (VII-VIII), 1962-2003.
- [18] E. Glaser, H. Christen, M. Friedli and M. Renn, *Kleiner Sprachatlas der deutschen Schweiz*, 5th ed., Frauenfeld/Stuttgart/Wien: Huber, 2013.
- [19] Phonogrammarchiv [Universität Zürich], *Der sprechende Atlas. "Gespräch am Neujahrstag" in 24 Dialekten*. Frauenfeld: Huber, 2000.
- [20] Praat. Doing phonetics by computer, 2015. www.praat.org, accessed 03 Mar 2015.
- [21] A. Leemann, M.-J. Kolly, and V. Dellwo, "Crowdsourcing regional variation in speaking rate through the iOS app 'Dialäkt App'," *Proc. Of Speech Prosody*, 2014.
- [22] A. Hanani, M. J. Russell and M. J. Carey, "Human and computer recognition of regional accents and ethnic groups from British English speech," *Computer Speech & Language*, vol. 27, pp. 59-74, 2013.
- [23] H. Behravan, V. Hautamäki and T. Kinnunen, "Foreign accent detection from spoken Finnish using i-vectors," *Proceedings of Interspeech 2013, Lyon*, pp. 79-83, 2013.
- [24] F. Biadsy, J. Hirschberg and M. Collins, "Dialect recognition using a phone-GMM-supervector-based SVM kernel," *Proceedings of Interspeech 2010 Makuhari*, pp. 753-756, 2010.
- [25] F. Biadsy, Automatic dialect and accent recognition and its application to speech recognition, PhD thesis, Columbia University, NY, 2011.
- [26] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore et al., "The HTK book (for HTK version 3.4)," Cambridge university engineering department, vol. 2, no. 2, pp. 2-3, 2006.
- [27] <http://www.pholab.uzh.ch/forschung/laufende-forschungsprojekte/VoiceApp/media.html>, accessed 03 Mar 2015.
- [28] M. Jessen, "Forensic reference data on articulation rate in German," *Science & Justice*, vol. 47, pp. 50-67, 2007.
- [29] H. Künzel, H. Masthoff and J. Köster, "The relation between speech tempo, loudness, and fundamental frequency: an important issue in forensic speaker recognition," *Science & Justice*, vol. 35, no. 4, pp. 291-295, 1995.
- [30] A. Leemann, M.-J. Kolly, R. Purves, D. Britain and E. Glaser, "Crowdsourcing language change with smartphone apps," *PLoS ONE*, in prep.
- [31] B. Laugwitz, "A web-experiment on colour harmony principles applied to computer user interface design," in U. D. Reips and M. Bosnjak (eds.), *Dimensions of Internet Science*. Germany: Pabst Science Publishers Lengerich, pp. 131-145, 2001.
- [32] R. Hotzenköcherle, *Einführung in den Sprachatlas der deutschen Schweiz*, Band A: *Zur Methodologie der Kleinraumatlantanten*, Bern: Francke, 1962.
- [33] M. Eskenazi, "An Overview," in: M. Eskenazi, G. A. Levow, H. Meng, G. Parent and D. Suendermann (eds.), *Crowdsourcing for Speech Processing: Applications to Data Collection, Transcription and Assessment*. UK: John Wiley & Sons, pp. 1-7, 2013.
- [34] M. H. Birnbaum, "Human research and data collection via the Internet," *Annu. Rev. Psychol.*, vol. 55, pp. 803-832, 2004.
- [35] U. D. Reips, "Standards for Internet-based experimenting," *Experimental psychology*, vol. 49, no. 4, p. 243, 2002.
- [36] I. McGraw, "Collecting speech from crowds," in M. Eskenazi, G. A. Levow, H. Meng, G. Parent and D. Suendermann (eds.) *Crowdsourcing for Speech Processing: Applications to Data Collection, Transcription and Assessment*. UK: John Wiley & Sons, 2013.
- [37] M. Cooke, J. Barker, M. L. Garcia Lecumberri, "Crowdsourcing for Speech Processing," in M. Eskenazi, G. A. Levow, H. Meng, G. Parent and D. Suendermann (eds.) *Crowdsourcing for Speech Processing: Applications to Data Collection, Transcription and Assessment*. UK: John Wiley & Sons, 2013.
- [38] M. Eskenazi, "The Basics," in M. Eskenazi, G. A. Levow, H. Meng, G. Parent and D. Suendermann (eds.) *Crowdsourcing for Speech Processing: Applications to Data Collection, Transcription and Assessment*. UK: John Wiley & Sons, 2013.
- [39] B. Juska-Bacher, C. Biemann and U. Quasthoff, "Webbasierte linguistische Forschung: Möglichkeiten und Begrenzungen beim Umgang mit Massendaten," *Linguistik online*, vol. 6, no. 4, 2013.
- [40] A. Leemann, M.-J. Kolly, I. Werlen, D. Britain and D. Studer-Joho, "The diffusion of /l/-vocalization in Swiss German," *Language Variation and Change*, vol. 26, no. 2, pp. 191-218, 2014.